Molecular Microbiology (2005) **57**(1), 212–221 doi:10.1111/j.1365-2958.2005.04670.x

Structural basis for segmental gene conversion in generation of *Anaplasma marginale* outer membrane protein variants

James E. Futse,[1] Kelly A. Brayton,[1] Donald P. Knowles, Jr[2] and Guy H. Palmer[1]*
[1]*Program in Vector-borne Diseases, Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA 99164-7040, USA.*
[2]*Animal Diseases Research Unit, USDA/ARS, Pullman, WA 99164-7030, USA.*

Summary

Bacterial pathogens in the genus *Anaplasma* generate surface coat variants by gene conversion of chromosomal pseudogenes into single-expression sites. These pseudogenes encode unique surface-exposed hypervariable regions flanked by conserved domains, which are identical to the expression site flanking domains. In addition, *Anaplasma marginale* generates variants by recombination of oligonucleotide segments derived from the pseudogenes into the existing expression site copy, resulting in a combinatorial increase in variant diversity. Using the *A. marginale* genome sequence to track the origin of sequences recombined into the *msp2* expression site, we demonstrated that the complexity of the expressed *msp2* increases during infection, reflecting a shift from recombination of the complete hypervariable region of a given pseudogene to complex mosaics with segments derived from hypervariable regions of different pseudogenes. Examination of the complete set of 1183 variants with segmental changes revealed that 99% could be explained by one of the recombination sites occurring in the conserved flanking domains and the other within the hypervariable region. Consequently, we propose an 'anchoring' model for segmental gene conversion whereby the conserved flanking sequences tightly align and anchor the expression site sequence to the pseudogene. Associated with the recombination sites were deletions, insertions and substitutions; however, these are a relatively minor contribution to variant generation as these occurred in less than 2% of the variants. Importantly, the anchoring model, which can account for more variants than a strict segmental sequence identity mechanism, is consistent with the number of *msp2* variants predicted and empirically identified during persistent infection.

Accepted 29 March, 2005. *For correspondence. E-mail gpalmer@vetmed.wsu.edu; Tel. (+1) 509 335 6033; Fax (+1) 509 335 8529.

Introduction

Persistent infection of mammalian hosts provides the opportunity for vector-borne pathogens to be acquired and transmitted by arthropod vectors. Persistence in immunocompetent hosts is commonly mediated by repeated generation of surface coat antigenic variants that evade the immune response (Borst, 1991; Deitsch *et al*., 1997; Barbour and Restrepo, 2000; Palmer *et al*., 2000). This is illustrated by, among others, protozoal pathogens in the genera *Plasmodium* and *Trypanosoma* (Barry and McCulloch, 2001; Turner, 2002) and by bacterial pathogens in the genera *Anaplasma* and *Borrelia* (Zhang *et al*., 1997; Zhang and Norris, 1998; Palmer *et al*., 2000; Brayton *et al*., 2001).

Unlike the relatively large genomes of protozoa, both *Anaplasma* and *Borrelia* have small bacterial genomes, <1.5 Mb, requiring efficiency in generation of a large number of variants (Fraser *et al*., 1997; Brayton *et al*., 2001). *Borrelia hermsii* and *B. burgdorferi*, causative agents of relapsing fever and Lyme disease, respectively, use similar mechanisms in which plasmid encoded genes encoding lipoproteins recombine into a single-expression site to generate surface coat variants (Zhang and Norris, 1998; Rich *et al*., 2001). In contrast, *Anaplasma marginale* and *A. phagocytophilum*, the causative agents of anaplasmosis in cattle and humans, respectively, do not contain plasmids and use chromosomally encoded pseudogenes as a template for recombination into single-expression sites via gene conversion (Brayton *et al*., 2001; 2002; Barbet *et al*., 2003). In *A. marginale*, the resulting major surface protein 2 (MSP2) variants express distinct surface domains, encode unique B and T cell epitopes, and evade the immune response to allow lifelong persistence in the mammalian reservoir host – critically important for ixodid ticks to acquire and transmit the pathogen (French *et al*., 1998; 1999; Palmer *et al*., 2000; Brown *et al*., 2003).

Although the complement of individual *A. marginale msp2* pseudogenes encoded within the chromosome (< 10 with differences among pathogen strains) is alone insufficient to generate the number of MSP2 variants actually expressed, the repertoire is greatly expanded by a related gene conversion mechanism in which an oligonucleotide segment of a pseudogene, rather than the whole hypervariable region of a given pseudogene, is recombined into the existing expression site *msp2* to create a unique variant (Brayton *et al*., 2002). This process, termed segmental gene conversion, uses multiple segments from each of the pseudogenes in a combinatorial fashion to generate the number of MSP2 variants required for long-term immune evasion and pathogen persistence.

The structural basis for gene conversion using the whole hypervariable region of a given *msp2* pseudogene is evident: the 5′ and 3′ sequences flanking the central hypervariable domains are identical among the pseudogenes and the expression site copy. However, the basis for recombination of the pseudogene segments is unknown and has significant consequences for generation of specific variants. Only specific pseudogene segments may be able to recombine into the expression site if identity with the existing expression site variant is required, and this could result in both an 'order' to the generation of segmental variants and some possible variants being 'disallowed' due to the lack of identity with any expression site variants. Examination of the structural basis for segmental gene conversion requires complete knowledge of the pseudogene sequences so that each expression site variant can be mapped to a specific pseudogene sequence. We have recently completed the genome sequence for the St. Maries strain of *A. marginale* (Brayton *et al*., 2005) and use this database to track the generation of variants during persistent infection. In this article, we report the testing of three linked hypotheses directed at understanding the structural basis for generation of *A. marginale* MSP2 variants. The first is that the complexity of MSP2 variants, defined as the number of segments derived from different pseudogenes, will increase with time during infection. This hypothesis presumes that recombination of whole hypervariable regions of pseudogenes is the most favoured event due to the identical flanking regions in the expression site copy (Brayton *et al*., 2001). Recombination of individual segments is predicted to be less favoured and, thus would not be apparent until the immune response provides the strong selective pressure necessary for selection of these less frequently generated variants. The second hypothesis to be tested is that recombination of individual segments is restricted by identity to the expression site variant sequence. A well-studied example of sequence identity-dependent recombination is illustrated in *Escherichia coli*, where a minimum identity of 20 base pairs is required for efficient recombination (Watt *et al*., 1985). This requirement for sequence identity with the expression site would be expected to result in favoured combinations of the individual segments and may result in a hierarchy or order in which specific variants can be generated. Finally, whether there are additional mechanisms for generating MSP2 variants such as mutations or deletions is unknown. Studies in *Borrelia*, *Neisseria*, *Trypanosoma* and *Plasmodium* have shown that multiple point mutations are frequently introduced into the previously expressed variable gene sequences to further diversify the antigenic repertoire (Lu *et al*., 1993; Restrepo and Barbour, 1994; Deitsch *et al*., 1997; Meyers *et al*., 2003). We examine this possibility for alternative mechanisms for generating *msp2* variants by testing a third hypothesis: that all the expression site variants can be accounted for by either whole or segmental pseudogene conversion.

Results

The generation of complex variants during acute and persistent infection

Successive populations of MSP2 variants that arose in each of four animals were evaluated during a 1-year period of infection following tick transmission of the St. Maries strain of *A. marginale*. The expression site variants were identified during acute bacteraemia (28–30 days following initiation of tick feeding) and at 2, 3, 4, 5, 6, 7, 8, 9 and 12 months starting from day one of the transmission tick feeding. At least 30 individual expression site variants were sequenced from each animal at each time point, for a total of 1333 individual sequenced *msp2* clones. The nucleotide and predicted amino acid sequences of each variant were mapped to the *A. marginale* St. Maries strain genome in order to assign the expressed oligonucleotide segments to specific chromosomal pseudogenes. The sequences of the five distinct pseudogenes were obtained from the complete genome sequence of the St. Maries strain (GenBank Accession No. CP000030).

To test whether the complexity of the expression site increased over time following tick transmission of the pathogen, each expression site hypervariable region sequence was individually aligned with the full complement of *msp2* pseudogenes (Fig. 1A). Figure 1B diagrams the multi-step gene conversion process that results in complex variants. We defined complexity in terms of the number of segmental changes derived from different pseudogenes, with 0 representing use of a whole hypervariable region of a given pseudogene, 1 representing a single segmental change, 2 representing two segmental changes from different pseudogenes, etc. The mean number of segmental changes during acute infection was 0.91 ± 0.11. After the resolution of the peak bacteraemia,

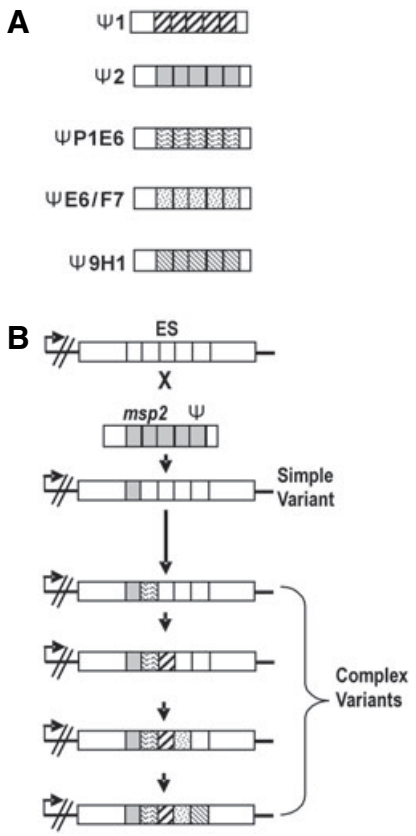

Fig. 1. Schematic representation of the complement of *msp2* pseudogenes and generation of complex variants.
A. *Anaplasma marginale* St. Maries strain contains five unique *msp2* pseudogenes (ψ). The unique central hypervariable regions (HVR) for individual pseudogenes are represented by different patterns, and are flanked by truncated domains encoding the conserved 5′ and 3′ regions, represented in white.
B. The single full-length *msp2* expression site (ES) HVR is shown as an open box, and the horizontal arrows indicate the direction of the promoter. Generation of complex variants involves multiple segmental changes to the expression site HVR. The resulting complex HVR sequences are mosaics not represented in any single pseudogene.

the mean number of segmental changes increased during persistent infection, reaching a mean of 2.58 ± 0.02 at 1 year after tick feeding (Fig. 2A). The increase in the mean number of segmental changes was significantly correlated with progression from acute bacteraemia to persistence ($r = 0.96$; $r^2 = 0.92$, $P = 0.03$). In addition, there was a significant increase ($P < 0.05$) in the frequency of complex variant types that require two or more recombination events into the expression site (≥ 2 individual segments) as compared with those achievable with a single recombination event. This is illustrated by the progression from the predominant recombination of whole hypervariable regions of pseudogenes and single pseudogene segments (generated using a single recombination event) during acute bacteraemia to predominance of two and three segmental changes at 1 year after infection (Fig. 2B).

Structural basis of segmental recombination

The overall sequence identity in the ~300-nucleotide hypervariable regions among the five unique *msp2* pseudogenes in the St. Maries strain is 39%. Between individual pairs of pseudogenes, the identity in the hypervariable region ranges from 41% between ψ9H1 and ψE6/F7 to 70% between ψ1 and ψE6/F7. This identity between pseudogene pairs is restricted to one to three blocks of 26–42 nucleotides with 61–98% identity. These blocks are shared between specific pairs of pseudogenes but differ distinctly from the remaining pseudogenes: each pseudogene shares at least one internal oligonucleotide block with an otherwise different pseudogene. Consequently, one model for segmental recombination would generate new expression site variants by identity within the hypervariable domains between a pseudogene segment and the existing expression site variant. This 'sequence identity' model (Fig. 3A) predicts that recombination would occur adjacent to the specific blocks of iden-

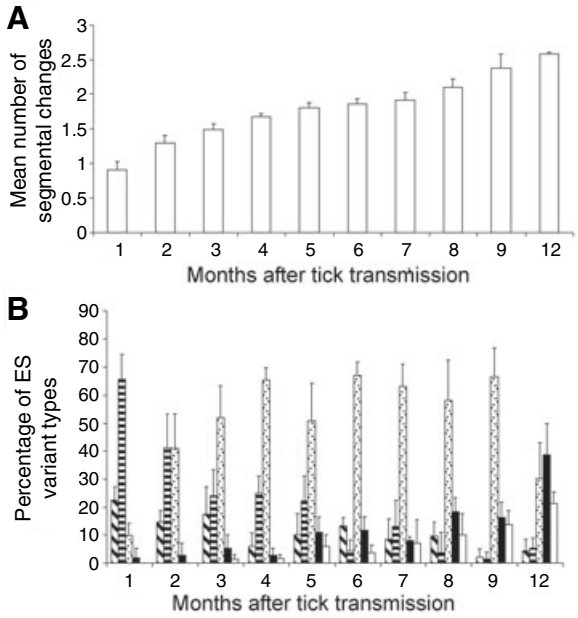

Fig. 2. Increasing complexity and frequency of complex variant types over time. The infection duration is shown on the *x*-axis.
A. The mean number of segmental changes in variants (*y*-axis) versus time during a 12-month period of infection.
B. The percentage (*y*-axis) of each variant type (whole hypervariable regions of pseudogenes; one to four different pseudogene segments) that arose at each time point: variants created by recombination of whole hypervariable regions of pseudogenes (▧) or by recombination of 1 (▤), 2 (▩), 3 (■) or 4 (□) oligonucleotide segments derived from different pseudogenes. The number of variants examined at each month was: month 1 (acute), 105; month 2, 130; month 3, 143; month 4, 136; month 5, 144; month 6, 144; month 7, 146; month 8, 140; month 9, 154; month 12, 91.

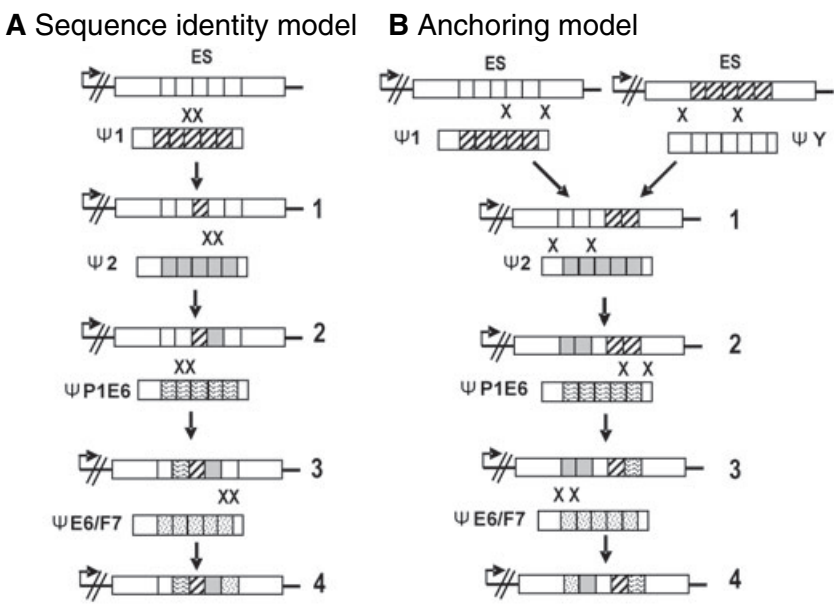

Fig. 3. Schematic representation of two models of segmental gene conversion. The expression site (ES) is diagrammed aligned with the pseudogenes (ψ). 'X' indicates the sites of recombination. 1, 2, 3 and 4 indicate the number of segmental changes derived from different pseudogenes.
A. This 'sequence identity' model predicts that segmental changes can initiate anywhere that identical nucleotides are shared between the expressed variant and the pseudogene. Some recombination events would not occur due to lack of identity between hypervariable region segments.
B. This 'anchoring' model predicts that the conserved, flanking 5′ and 3′ domains tightly align and anchor the expression site sequence to the pseudogene. One recombination site will occur in the conserved 5′ or 3′ end and the other recombination site can occur at any point along the HVR; sequence identity is not required for this second recombination event.

tity within the hypervariable regions that are shared among specific pseudogene pairs. Thus, new variants would have segments recombined into the 5′, middle or 3′ blocks of existing hypervariable regions with the complex variants observed in persistent infection being generated by sequential segmental recombination events at any block with sufficient identity in the expression site copy. Alternatively, a second model for segmental recombination presumes that identity in the conserved 5′ and 3′ domains anchors the two members of the recombination duplex tightly together. One site of recombination falls within one of the conserved ends, and the other site of recombination falls within the hypervariable region, and this second site does not require sequence identity (Fig. 3B). In this 'anchoring' model, no segmental recombination events would be expected within the internal or middle blocks of the existing expression site hypervariable domain.

Of the 1333 variants examined, 150 resulted from recombination of whole hypervariable regions of pseudogenes, 282 from segments of a single pseudogene and 901 from recombination of two or more segments derived from different pseudogenes. All 282 expression site variants containing single segmental changes could be assigned to the block immediately flanking the highly conserved 5′ or 3′ domains. Examination of the 901 expression site variants with multiple segmental changes revealed that 888 contained segmental changes immediately downstream or upstream from the conserved 5′ or 3′ regions, respectively, whereas only 13 variants contained segmental changes in the middle. This bias towards the 5′ and 3′ regions as compared with the middle blocks of the hypervariable region was statistically significant at $P < 0.05$ (χ^2 test of likelihood ratio). Finally, examination of adjacent segments in the complete set of 1183 variants with hypervariable regions from different pseudogenes revealed overlapping identity of > 3 and > 20 nucleotides in only 8.9% and 2.4% respectively.

The two recombination models also differ in that only the 'sequence identity' model would confer an order or restriction upon generation of segmental variants based on identity between the hypervariable domains of the individual pseudogenes. To search for an ordered generation of segmental variants, we examined whether the same variants arose in each of the four infected animals and if so, whether they arose at similar time points during persistent infection. Each of the unique whole hypervariable regions derived from the five pseudogenes could be identified within the expression site in each animal. However, analysis of 152 unique variants revealed that only 43 occurred in two animals and only eight occurred in three animals. No segmental variant was identified in all four animals. Of the 51 segmental variants common to at least two animals, none were found at the same time point of infection.

Identification of additional mechanisms for variant generation

For 1317/1333 expression site variants examined, the complete sequence could be mapped to either a complete pseudogene hypervariable region or, one or more segments from within these hypervariable regions. Six expression site variants had nucleotide deletions at the junctions where a segment of one pseudogene had recombined with a different pseudogene-derived

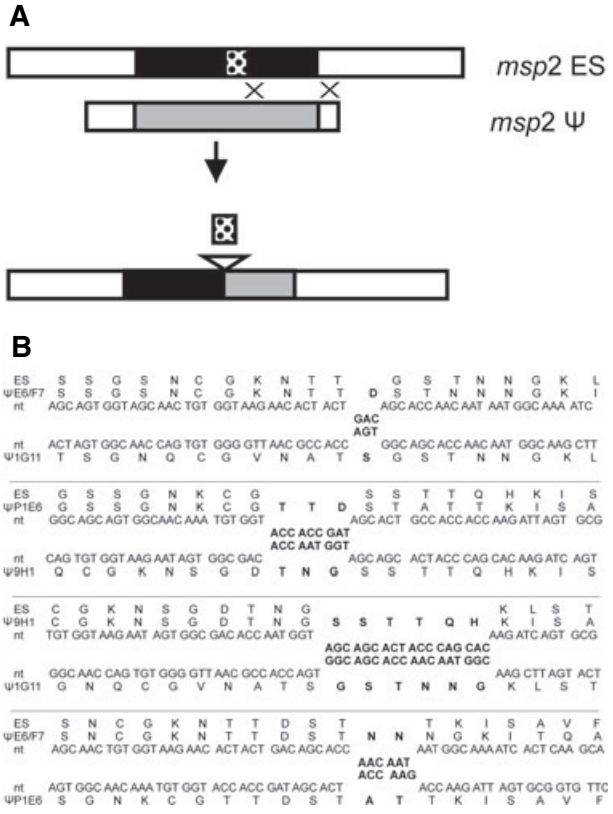

Fig. 4. Occurrence of deletions at segmental switch sites.
A. Schematic representation of recombination of a pseudogene derived segment (grey) into the 3′ end of the existing expression site (ES) variant (black). Recombination sites are indicated by 'X' and the deleted sequence is indicated by the stippled box (⚄).
B. Switch sites of four individual variants. In each panel, ES and ψ represent the expression site and pseudogene sequences, respectively, immediately downstream from the site of the most 5′ segmental change. The deleted nucleotides (nt) and the encoded amino acids are shown in bold face.

sequence within the expression site. Of these six, three were detected at acute bacteraemia in animals 990 and 983. The remaining three variants arose at different time points during persistent infection in animals 983, 985 and 995. The shortest deleted segments observed were three nucleotides long, and the longest were 18 nucleotides (Fig. 4). Of the four different deletion types, only one was flanked by a short 3 bp direct repeat.

In addition to the deletion variants, 10 expression site variants contained sequences that could not be mapped to an individual pseudogene. Seven of these variants had discrete insertions within the hypervariable region at the junctions of segments derived from different pseudogenes. These insertion variants encoded either STNQ or TATSSTSTG oligopeptide segments; the STNQ variant was detected in animals 990 and 995 during acute and persistent bacteraemia, respectively, and the larger insertion was detected during persistence in animals 983 and 995. Neither insertion could be mapped to the St. Maries strain genome as either a linear nucleotide sequence encoding the complete insertion or as smaller, nine-nucleotide segments that, if recombined in a novel mosaic, would generate the insertion. Three expression site variants, derived from acute and persistent infection, respectively, in animals 990 and 995, contained a block of sequence in the first 5′ segment of the hypervariable region that was most similar to, but not identical with, the corresponding segment in pseudogene P1/E6. In contrast to those variants with discrete insertions or deletions, these three variants had multiple substitutions (11 in the first 72 nt of the hypervariable region) relative to ψP1/E6, resulting in six-amino-acid substitutions (Fig. 5A). To examine whether the substitutions were also present in a pseudogene locus, reflecting either a change in ψP1/E6 or emergence of a separate clone within the St. Maries strain with a unique pseudogene, Southern blot analyses of unamplified genomic DNA were performed using a digoxigenin-labelled pan-*msp2* probe and a 27 bp oligonucleotide probe derived from the expressed substitution variant sequence (designated the KAV sequence based on the N→K amino acid change in the variant relative to ψP1/E6). Using DNA derived from the time point when the KAV variant was expressed in animal 990, the 27 bp KAV probe hybridized only with the 3.7 kb *Fsp*I fragment (Fig. 5B), which corresponds to the *msp2* expression site (Barbet *et al*., 2000). In contrast, the pan-*msp2* probe hybridized to additional fragments representing the full complement of the pseudogenes (Fig. 5B) as well as the expression site.

Discussion

Gene conversion of either a whole hypervariable region or a single segment into the *msp2* expression site can be achieved in a single recombination event. Consequently, these can be classified as 'simple variants', while those with two or more segments requiring multiple recombination events represent 'complex variants' (Fig. 1B). Our data demonstrate that simple variants arise early in infection, forming the majority of variants during acute bacteraemia. This predominance of simple variants most likely represents a significantly higher frequency of generation with emergence of complex variants only when the immune system has controlled the high level acute bacteraemia and provides a strong selective pressure for additional variants requiring two or more recombination events. Consistent with immune selection for the complex variants, the number of simple variants decreased dramatically over time (Fig. 2). An alternative explanation for a higher frequency of generation of the simple variants is an essentially equal frequency of generation for both types of variants with the simple variants having a replication or growth rate advantage *in vivo*. Thus, simple type

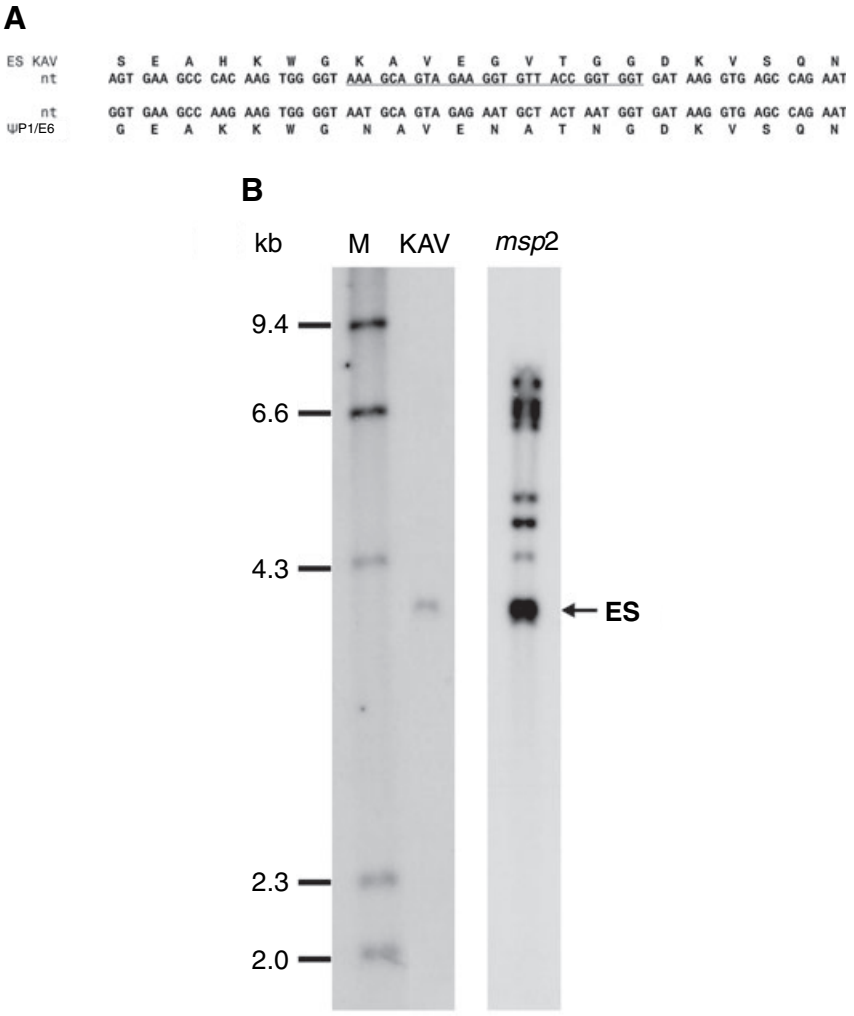

Fig. 5. Substitutions specific to the KAV expression site variant.
A. Sequence of the 5′ segment of the KAV expression site variant, detected during infection in animals 990 and 995, aligned with the corresponding segment in the most closely related pseudogene, ψP1/E6. The sequence used as the KAV variant probe is underlined.
B. Southern blot analysis of St. Maries strain genomic DNA isolated from animal 990 at the time of expression of the KAV variant. The left panel was hybridized with the KAV-specific oligonucleotide probe (underlined sequence in A). The right panel shows the same membrane hybridized with the pan-*msp2* probe after stripping and re-probing. M designates the molecular size markers; ES designates the position of the expression site.

variants would predominate early in acute bacteraemia before the onset of the immune response. However, direct inoculation of blood from persistently infected animals into naive animals or splenectomy of persistently infected animals results in high level *A. marginale* bacteraemia, suggesting that the variants arising in persistent infection are capable of replicating to the same high levels (10^9 per ml of blood) as simple variants (Kieser *et al*., 1990; Barbet *et al*., 2001).

The first hypothesis that complexity of MSP2 variants increases with time during infection is accepted. Increasing complexity results in progressive generation of expression site variants that are mosaics of different pseudogene sequences. The identity in hypervariable region blocks between specific pairs of pseudogenes, combined with the requirements for efficient recombination established using *E. coli* (Watt *et al*., 1985), led us to hypothesize that recombination of individual segments is restricted by identity to the expression site variant sequence and, thus, that certain segmental gene conversion events would be favoured. We reject this second hypothesis based on the following lines of evidence: (i) identity of more than three nucleotides flanking the junction between the donor hypervariable region sequence and the recipient expression site hypervariable region sequence was detected in only 8.9% of the segmental recombination events and identity of more than 20 nucleotides in only 2.4%, (ii) all 282 single segmental changes could be explained by recombination of pseudogene sequence at either end of the expression site hypervariable region, adjacent to the highly conserved 5′ and 3′ flanking domains, (iii) 888/901 variants with multiple segmental changes could be explained by recombination of pseudogene sequence adjacent to the conserved 5′ and 3′ flanking domains with only 13 being derived from recombination of a donor hypervariable region segment into an internal block of the expression site hypervariable domain and (iv) there was no preferential recombination between segments of the most closely related pseudogenes. In the absence of any favoured segmental recombination events, an order to

segmental variant generation would be unlikely. Consistent with a lack of ordered progression, emergence of identical segmental variants among individual persistently infected animals was uncommon with no evidence of a temporal pattern.

The process of homologous recombination requires DNA strand breaks followed by homologous pairing, heteroduplex extension and ultimately mismatch repair (Meselson and Radding, 1975). As a consequence of mismatch repair, deletions, insertions and substitutions can be generated at the recombination sites. We identified deletions (Fig. 4) as well as insertions of novel sequence but only at the junctions between oligonucleotide segments. The inserted sequence could not be mapped to any existing *A. marginale* sequence, including the pseudogenes, by analysis of the genome sequence. We considered the possibility that the insertions could have been derived from an alternative genomic template. However, neither insertion could be mapped to the St. Maries strain genome as either a linear nucleotide sequence encoding the complete insertion or as smaller, nine-nucleotide pseudogene segments that, if recombined in a novel mosaic, would generate the insertion. Additionally, a variant with multiple substitutions in the most 5′ segment of the hypervariable region, adjacent to the recombination site created by the identity in the 5′ conserved domain, was identified. Similar to the insertion variants, this sequence was not represented in the genome sequence and the Southern blot analysis excluded the possibility that a clone within the St. Maries strain contained a pseudogene with this alternative sequence. That the variant sequence was not present in the genome aside from the expression site, combined with the position of the substitutions adjacent to recombination sites, is consistent with, but not definitive evidence of, their generation during recombination. Amino acid substitutions in the 5′ segment of the hypervariable region have been shown to result in loss of T cell recognition (Brown *et al*., 2003) and thus, substitution variants would have a selective advantage in an infected immunocompetent host. Consequently, the third hypothesis that all the expression site variants can be accounted for by either whole or segmental pseudogene conversion is rejected. However, the frequency of detection of these alternatively generated variants was very low, accounting for 1.2% of all variants. Thus, these mechanisms of deletion, insertion and substitution do not appear to be major contributors to MSP2 variation in persistent *A. marginale* infection. Whether the low frequency of detection reflects the actual frequency of generation is unknown. Deletions or insertions that alter the reading frame or substitutions that introduce a stop codon resulting in a truncated protein or changes that alter the carboxy-terminal domain, which is highly conserved and apparently required, would presumably not be detected due to failure of the organism to invade, replicate or survive in the host.

The data presented support the ‘anchoring’ model of recombination (Fig. 6) as the preferred mechanism for the generation of complex variants in *msp*2. Unlike *E. coli* (Watt *et al*., 1985), there is no apparent minimum region of sequence identity required for recombination in the hypervariable region; however, this is only possible due to the strong anchoring of the conserved ends holding the two recombining molecules in close proximity. Structurally, this results in segmental changes occurring adjacent to one or the other of the conserved ends. Functionally, this results in the generation of more variants than can be created if a strict sequence identity mechanism was employed. Coupled with the less favoured alternative methods of generating sequence variation (insertions, deletions and substitutions following mismatch repair, and segmental recombination at sites of sequence identity), a large number of complex variants can be generated over time even with a very limited number of pseudogenes. This is strongly supported by the detection of 26–43 variants per animal over the 1-year course of infection, consistent with the predicted number required for evasion of the immune response (Palmer *et al*., 2000; Brayton *et al*., 2002).

Experimental procedures

Pathogen genome

The St. Maries strain of *A. marginale* was used in all experiments as the complete genome sequence, required for mapping expression site sequences, was available (Brayton *et al*., 2005). The *msp2* gene complement consists of a single operon-linked expression site, and seven pseudogenes. However, two pseudogenes are duplicated resulting in only five unique pseudogene sequences (Fig. 1A). Pseudogene (ψ) 1 and 2 were previously reported to GenBank (Accession No. AF305077). The remaining pseudogenes were obtained from the complete genome sequence (GenBank Accession No. CP000030).

Transmission feeding on naive calves

Dermacentor species mediate tick transmission of *A. marginale* strains in the temperate regions of the USA (Stiller and Coan, 1995). Males from the Reynolds Creek stock of *Dermacentor andersoni* (free of *A. marginale* and other pathogens) were selected from engorged nymphs that had molted to the adult stage following incubation for 48 h at 26°C and 92% relative humidity. Tick transmission of the St. Maries strain to cattle using the Reynolds Creek *D. andersoni* has been previously reported (Eriks *et al*., 1993; Futse *et al*., 2003). Holstein calves (3–4 months of age), confirmed to be *A. marginale* negative by the MSP-5 CI-ELISA (Torioni de Echaide *et al*., 1998), were each infected by tick transmission with the St. Maries strain. A group of 30 *D. andersoni* adult

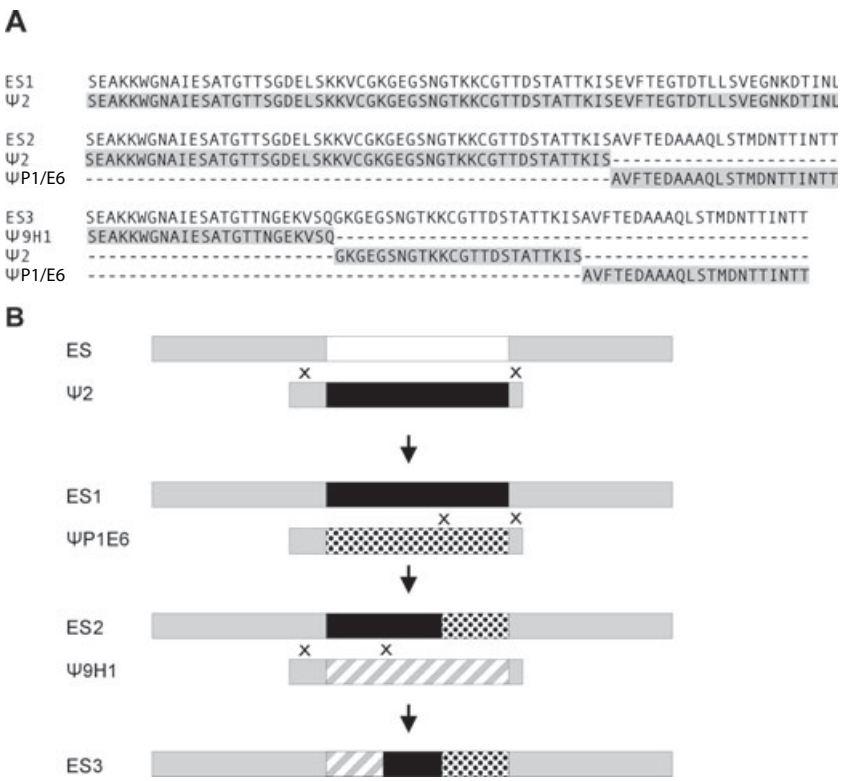

Fig. 6. Generation of complex mosaics by the anchoring model of gene conversion. The existing expression site sequence (ES1) represents the whole hypervariable region of pseudogene 2. Recombination of a segment of pseudogene P1/E6 into the 3′ end of the ES1 hypervariable region generates ES2, representing a single segmental variant. Recombination of a segment of pseudogene 9H1 into the 5′ end of the ES2 hypervariable region generates ES3, a complex variant derived from three different pseudogenes.
A. Hypervariable region amino acid sequences of ES variants 1–3. The pseudogene sequence (shaded) that serves as the template is provided below each ES variant.
B. Schematic representation of the recombined segments and recombination sites (indicated by 'X').

males previously infected with the St. Maries strain were allowed to transmission feed on each of four individual Holstein calves (983, 985, 990, 995) for 6 days. During acute infection, Giemsa-stained blood smears of individual recipient calves were examined daily for microscopically detectable *A. marginale*. Serum samples were monitored weekly using CI-ELISA to confirm seroconversion. Calves were considered persistently infected after no infected erythrocytes were observed by light microscopy for 3 consecutive weeks, consistent with a bacteraemia level of $\leq 10^7$ organisms per ml of infected blood (Kieser *et al*., 1990). The actual *A. marginale* levels of individual infected calves were determined by quantitative real-time polymerase chain reaction (PCR) using Taqman assay for *msp5* (Futse *et al*., 2003).

Isolation, amplification, cloning and sequencing of msp2 *expression site DNA*

Total *A. marginale* genomic DNA was extracted from peripheral blood samples from each of four calves using the Puregene isolation protocol (Gentra Systems). Blood samples were taken from acute bacteraemia through 12 months for calves 983, 990 and 995 and from 2 to 10 months after infection for calf 985. The expression site *msp2* hypervariable region was amplified from genomic DNA templates as previously described (Brayton *et al*., 2001) using forward (5′-TCC TAC CAA GCG TCT TTT CCC C-3′) and reverse (5′-TTA CCA CCG ATA CCA GCA CAA-3′) primers that specifically amplify the hypervariable region of the *msp2* expression site. Standard PCR amplifications were performed in reaction volumes of 25 μl using 0.5 μM of each primer. Amplification reactions were performed in a Perkin Elmer GeneAmp® PCR thermocycler (Perkin Elmer. Norwalk, CT) with an initial denaturation step of 95°C for 5 min, followed by 30 cycles of 94°C for 20 s, 60°C for 20 s and 72°C for 1 min. The products were extended at 72°C for 7 min before holding the reaction at 10°C. The amplicons were size-separated and visualized in a 1% agarose gel electrophoresis following staining with ethidium bromide. The expression site hypervariable region amplicons were cloned into the PCR-4 TOPO vector using the TOPO-TA cloning kit (Invitrogen) and transformed into TOP10 *E. coli* cells. Plasmid DNA was isolated from the transformed colonies using Wizard plus Miniprep DNA purification system (Promega, Madison, WI), and the presence of inserts was confirmed by *Eco*RI digestion. Plasmid inserts were sequenced in both directions using the Big Dye kit and an ABI Prism automated sequencer (Applied Biosystems).

Southern blots and hybridization

Total genomic DNA from animal 990 at a time point containing the substitution KAV variant was digested with the restriction endonuclease *Fsp*I, and the fragments were size-separated by 0.8% agarose gel electrophoresis and transferred to a nylon membrane. The membrane was hybridized with a digoxigenin-labelled oligonucleotide probe derived from the substitution KAV variant sequence (AAA GCA GTA GAA GGT GTT ACC GGT GGT), and subsequently, with a pan-*msp2* probe corresponding to nucleotides 2–965 of *msp2* (numbering according to Palmer *et al*., 1994). Oligonucleotide hybridization was as follows: pre-hybridization at 57.3°C in a buffer containing 5× SSC, 1× blocking buffer

(Roche Molecular Biochemicals), 0.1% sarcosine (w/v), 0.02% SDS (w/v) and 0.1 mg ml^{-1} poly-(A) for 2 h. The buffer was replaced with fresh buffer containing 20 ng of the KAV probe and hybridized overnight at the same temperature. The membrane was washed twice in 2× SSC, 0.1% SDS at 72°C for 5 min followed by two washes in 0.1× SSC, 0.1% SDS at 72°C for 10 min. Following hybridization and detection with the KAV variant-specific probe, the same membrane was stripped and rehybridized with the pan-*msp2* probe at 42°C in Dig EasyHyb. Generation of the probe, pre-hybridization, hybridization, detection and stripping were performed as recommended by the manufacturer (Roche Molecular Biochemicals). The membrane was washed twice in 2× SSC, 0.1% SDS at room temperature for 15 min, followed by two 15 min washes in 2× SSC, 0.1% SDS at 65°C, and 0.2× SSC, 0.1% SDS at 65°C.

Acknowledgements

This research was supported by NIH R01 AI44005, USDA-ARS-CRIS 5348-32000-016-00D and USDA-ARS Cooperative Agreement 58-5348-3-212. The technical assistance of Ralph Horn and Bev Hunter is gratefully acknowledged.

References

Barbet, A.F., Lundgren, A., Yi, J., Rurangirwa, F.R., and Palmer, G.H. (2000) Antigenic variation of *Anaplasma marginale* by expression of MSP2 mosaics. *Infect Immun* **68:** 6133–6138.

Barbet, A.F., Yi, J., Lundgren, A., McEwen, B.R., Blouin, E.F., and Kocan, K.M. (2001) Antigenic variation of *Anaplasma marginale*: major surface protein 2 diversity during cyclic transmission between ticks and cattle. *Infect Immun* **69:** 3057–3066.

Barbet, A.F., Meeus, P.F., Belanger, M., Bowie, M.V., Yi, J., Lundgren, A.M., *et al.* (2003) Expression of multiple outer membrane protein sequence variants from a single genomic locus of *Anaplasma phagocytophilum*. *Infect Immun* **71:** 1706–1718.

Barbour, A.G., and Restrepo, B.I. (2000) Antigenic variation in vector-borne pathogens. *Emerg Infect Dis* **6:** 449–457.

Barry, J.D., and MacCulloch, R. (2001) Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv Parasitol* **49:** 1–70.

Borst, P. (1991) Molecular genetics of antigenic variation. *Immunol Today* **12:** A29–A33.

Brayton, K.A., Knowles, D.P., McGuire, T.C., and Palmer, G.H. (2001) Efficient use of small genome to generate antigenic diversity in tick-borne ehrlichial pathogens. *Proc Natl Acad Sci USA* **98:** 4130–4135.

Brayton, K.A., Palmer, G.H., Lundgren, A., Yi, J., and Barbet, A.F. (2002) Antigenic variation of *Anaplasma marginale msp2* occurs by combinatorial gene conversion. *Mol Microbiol* **43:** 1151–1159.

Brayton, K.A., Kappmeyer, L.S., Herndon, D.R., Dark, M.J., Tibbals, D.L., Palmer, G.H., *et al.* (2005) Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc Natl Acad Sci USA* **102:** 844–849.

Brown, W.C., Brayton, K.A., Styer, C.M., and Palmer, G.H. (2003) The hypervariable region of *Anaplasma marginale* major surface protein 2 (MSP2) contains multiple immunodominant CD4+ T lymphocyte epitopes that elicit variant-specific proliferative and IFN-γ responses in MSP2 vaccinates. *J Immunol* **170:** 3790–3798.

Deitsch, K.W., Moxon, E.R., and Wellems, T.E. (1997) Shared themes of antigenic variation and virulence in bacterial, protozoal and fungal infections. *Microbiol Mol Biol Rev* **61:** 281–293.

Eriks, I.S., Stiller, D., and Palmer, G.H. (1993) Impact of persistent *Anaplasma marginale* rickettsemia on tick infection and transmission. *J Clin Microbiol* **31:** 2091–2096.

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.R., Lathigra, O., *et al.* (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **6660:** 553–555.

French, D.M., McElwain, T.F., McGuire, T.C., Brown, W.C., and Palmer, G.H. (1998) Expression of *Anaplasma marginale* major surface protein 2 variants during persistent cyclic rickettsemia. *Infect Immun* **66:** 1200–1207. Erratum, *Infect Immun* **66:** 2400.

French, D.M., Brown, W.C., and Palmer. G.H. (1999) Emergence of *Anaplasma marginale* antigenic variants during persistent rickettsemia. *Infect Immun* **67:** 5834–5840.

Futse, J.E., Ueti, M.W., Knowles, D.P., and Palmer, G.H. (2003) Transmission of *Anaplasma marginale* by *Boophilus microplus*: retention of vector competence in the absence of vector-pathogen interaction. *J Clin Microbiol* **41:** 3829–3824. Erratum, *J Clin Microbiol* **41:** 5354.

Kieser, S.T., Eriks, I.S., and Palmer, G.H. (1990) Cyclic rickettsemia during persistent infection of cattle with *Anaplasma marginale*. *Infect Immun* **58:** 1117–1119.

Lu, Y., Hall, T., Gay, L.S., and Donelson, J.E. (1993) Point mutations are associated with a gene duplication leading to the bloodstream re-expression of a trypanosome metacyclic VSG. *Cell* **3:** 397–406.

Meselson, M., and Radding, C.M. (1975) A general model for genetic recombination. *Proc Natl Acad Sci USA* **72:** 358–361.

Meyers, L.A., Levin, B.R., and Richardson, A.R. (2003) Epidemiology, hypermutation, within-host evolution, and virulence of *Neisseria meningitidis*. *Proc R Soc Lond* **270:** 1667–1677.

Palmer, G.H., Eid, G., Barbet, A.F., McGuire, T.C., and McElwain, T.F. (1994) Immunoprotective *Anaplasma marginale* major surface protein 2 is encoded by a polymorphic multigene family. *Infect Immun* **62:** 3808–3816.

Palmer, G.H., Brown, W.C., and Rurangirwa, F.R. (2000) Antigenic variation in the persistence and transmission of the ehrlichia *Anaplasma marginale*. *Microbes Infect* **2:** 167–176.

Restrepo, B.I., and Barbour, A.G. (1994) Antigenic diversity in the bacterium *B. hermsii* through 'somatic' hypermutations in rearranged *vmp* genes. *Cell* **5:** 867–876.

Rich, S.M., Sawyer, S.A., and Barbour, A.G. (2001) Antigen polymorphism in *Borrelia hermsii*, a clonal pathogenic bacterium. *Proc Natl Acad Sci USA* **26:** 15038–15043.

Stiller, D., and Coan, M.E. (1995) Recent developments in elucidating tick vector relationships for anaplasmosis. *Vet Parasitol* **57:** 97–108.

Torioni de Echaide, S., Knowles, D.P., McGuire, T.C., Palmer, G.H., Suarez, C.E., and McElwain, T.F. (1998) Detection of cattle naturally infected with *Anaplasma marginale* in a region of endemicity by nested PCR and a competitive enzyme-linked immunosorbent assay with recombinant major surface protein 5. *J Clin Microbiol* **36:** 777–782.

Turner, C.M.R. (2002) A perspective on clonal phenotypic (antigenic) variation in protozoan parasites. *Parasitology* **125:** S17–S23.

Watt, V.M., Ingles, C.J., Urdea, M.S., and Rutte, W.J. (1985) Homology requirements for recombination in *Escherichia coli*. *Proc Natl Acad Sci USA* **82:** 4768–4772.

Zhang, J.R., and Norris, S.J. (1998) Genetic variation of the *Borrelia burgdorferi* gene *vls* involves cassette-specific segmental gene conversion. *Infect Immun* **66:** 3698–3702.

Zhang, J.R., Hardham, J.M., Barbour, A.G., and Norris, S.J. (1997) Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89:** 275–285.